Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 586236, 18 pages
http://dx.doi.org/10.1155/2014/586236

Research Article

Service Capacity Reserve under Uncertainty by Hospital's ER Analogies: A Practical Model for Car Services

Miguel Ángel Pérez Salaverría[1] and José Manuel Mira McWilliams[2]

[1] *Jaguar Land Rover SLU, Torre Picasso, Plaza Pablo Ruiz Picasso 1, Planta 42, Complejo Azca, 28020 Madrid, Spain*
[2] *Universidad Politécnica de Madrid, Avenida Ramiro de Maeztu 7, 28040 Madrid, Spain*

Correspondence should be addressed to José Manuel Mira McWilliams; josemanuel.mira@upm.es

Received 23 January 2014; Revised 19 May 2014; Accepted 21 May 2014; Published 10 July 2014

Academic Editor: Michael Lütjen

We define a capacity reserve model to dimension passenger car service installations according to the demographic distribution of the area to be serviced by using hospital's emergency room analogies. Usually, service facilities are designed applying empirical methods, but customers arrive under uncertain conditions not included in the original estimations, and there is a gap between customer's real demand and the service's capacity. Our research establishes a valid methodology and covers the absence of recent researches and the lack of statistical techniques implementation, integrating demand uncertainty in a unique model built in stages by implementing ARIMA forecasting, queuing theory, and Monte Carlo simulation to optimize the service capacity and occupancy, minimizing the implicit cost of the capacity that must be reserved to service unexpected customers. Our model has proved to be a useful tool for optimal decision making under uncertainty integrating the prediction of the cost implicit in the reserve capacity to serve unexpected demand and defining a set of new process indicators, such us capacity, occupancy, and cost of capacity reserve never studied before. The new indicators are intended to optimize the service operation. This set of new indicators could be implemented in the information systems used in the passenger car services.

1. Introduction and Literature Review

Today, the passenger car industry is one of the world's most important industries encompassing investment groups and manufacturers. All passenger car brands operate in a global competitive marketplace with commercial brands that must offer a wide range of products, including repair and maintenance services.

Historically, passenger car services were intended to fix product issues and carry out the scheduled maintenance routines. However, at present time, after-sale services have evolved, becoming an indispensable part of the business to ensure current customer retention and new customer conquest. The after-sales service market has ballooned to four to five times the size of the original equipment business [1].

Under the above scope, any after-sales service opportunity is taken into account, not only to fix or maintain the car but also to respond to customer demands and increase company's revenue. Escalating customer expectations for rapid, flawless service support has increased the opportunity for firms to profit from appropriately priced differentiated service products targeted to meet the needs of particular market segments [1].

Thus, customer demands are not exclusively related to product issues; therefore, services are conveniently designed to suit customer needs and exceed initial expectations to make sure clients remain loyal to the brand and keep purchasing new products. On the other hand, services are usually planned in advance and customers are required to arrange an appointment prior to visiting the workshop, but whenever there is a breakdown, servicing unexpected visits introduces a random component and its resolution will always depend on the workshop availability.

As a consequence of the additional challenges in the after sales we have described in the previous paragraphs that passenger car companies embrace commercial relationships with a focus on maximizing revenue. That revenue will be obtained only if the direct result of the customer lifetime value is positive. With this in mind, passenger car brands adopt the motto "a happy customer is a returning customer," but there is an important difference between the meanings of satisfaction and retention [2]. Nevertheless, even when everything has been carefully planned, an unexpected customer might appear, and independently of how the service manages the customer, the emergency will affect the service revenue.

A very recent research [3] has demonstrated that the integrated nature of the after-sales quality in the passenger car service is strongly associated with the retention rate of the customers. In that study, the authors confirmed that when customers perceive the poor service quality, immediately they switch to another service centre. In this study it is also proved that, in this highly competitive environment, it is the service quality only by which brands can retain their customers. This confirms there is a real gap [4] between customer's real demand and actual service capacity.

The above gap is also studied by other authors. Literally citing an article published by Cohen et al. [1] in 2006, "Customers don't expect products to be perfect, but they do expect manufacturers to fix things quickly when they break down. Not surprisingly, customers are usually unhappy with the quality of after-sales support." According to the same publication, "That's mainly because after-sales support is notoriously difficult to manage, and only companies that provide services efficiently can make money from them."

Essentially, passenger car breakdowns are unexpected and do not adhere to planned schedules, like maintenance. Only those passenger car brands that manage after-sales service skilfully make money from it [1, 5].

1.1. Service Quality and Service Capacity. Service quality as a generic concept is well defined by various researchers in several ways [6–9]; technical quality, functional quality, and reputation are identified as the most frequent components of service quality. Usually, passenger car brands measure service quality by comparing initial customer expectations before the service with the perception after it has been delivered.

While service quality is a popular term in the passenger car industry, service capacity is not. Only limited research has been published on service capacity in the passenger car industry with respect to the extra capacity required to serve unexpected demand. Recent service capacity studies focus mainly on the specific situation in emerging markets, such as China and India, but no new researches have been published in regard to mature markets, such as Europe or USA.

Although there are no new specific publications in the passenger car field from an operations research approach, there are other studies from a marketing perspective [3, 10, 11]. This means that the issue of service capacity in the passenger car service industry has hardly been dealt with [7].

1.1.1. A Very Different Approach: Hospital's Service Capacity. As opposed to the situation in the passenger car industry, hospitals often reserve capacity for patients arriving to their hospital's emergency room (ER) in response to demand uncertainty. Reserving part of the hospital's capacity ensures enough flexibility for urgent admissions. Particularly, this is the usual scenario for premium passenger car manufacturers and traders, but it is not limited to them. Passenger car brands set up processes to ensure that customers are taken care of with the maximum convenience, which allows us to compare workshops with hospital's emergency room (ER). Under this scope, workshop bays are intended as ER beds, technicians are like the medical staff and the service reception and foreman must act like the hospital's emergency room (ER) capacity planner. Thus, there is a tradeoff to be optimized between service efficiency and the capability to admit unexpected customers in the process by reserving some of the service capacity.

The seminal references in the healthcare sector for the present document are based on the work of Kamenetzky et al. [12], who studied, in 1982, how to estimate necessities and demands for prehospital care. Subsequently in 1993, Badri and Hollingsworth [13] published a simulation model for scheduling in the hospital's emergency room (ER). Later, in 1996, Gerchak et al. [14] studied a reservation planning under uncertain demand for emergency surgery. Additionally, in 1998, Bazargan et al. [15] set an initial approach to hospital's emergency room (ER) and hospital services utilization using a theoretical model from historical data (kind of patient, demography, etc.).

Other authors approached hospital's capacity problem from an operational research point of view. In 2004, Brailsford et al. [16] developed a model for emergency and on-demand health care for large complex systems. Also in 2004, Beraldi et al. [17] created a stochastic programming routine with probabilistic constraints aimed to solve a location and dimensioning problem. Then in 2006, Green et al. [18] developed a model to manage patient service in a diagnostic medical facility.

Unlike the passenger car industry, hospitals usually manage their hospital's emergency room (ER) capacity by making the distinction between elective and emergency (nonelective) admissions and highlight the importance of an accurate forecast on both [19], estimating how fixed capacity on the nonelective admission expectations of unexpected demand turns into effective demand.

In other words, if a hospital capacity requires a number of available beds to assign to incoming patients, health managers have capacity for those patients who might enter the hospital as elective demand; that is, after a specialist diagnoses and retains some of the whole hospital capacity for those patients entering from the emergency room, because as depending on the severity of their disease they might not be rejected.

In contrast to what is described as a standard process in the passenger car arena, hospitals usually reserve some capacity in response to demand uncertainty to support the specification for optimal capability, incorporating cost derived from capacity reservation. By reserving some "empty" beds in the hospital, capacity planners ensure the required capacity to serve "emergency" admissions of patients.

Unfortunately, the literature review, in regard to health management, confirms that seldom estimations of hospital cost structures have taken production into account by incorporating the impact of nonelective demand on hospital cost structures. The same papers establish that hospitals are in control of the output decisions, in response to such unexpected demand [20]. In these studies the emphasis has been on estimating (and minimizing) the cost of maintaining reserve capacity rather than using nonelective demand as part of a decision support system. Our research will lean on the work referred to in this paragraph to apply the proposed methodology to the unexpected demand in the passenger car service industry.

1.2. Usual Tools and Models in the Passenger Car Industry and the Service Sector. An extended tool along the passenger car industry is an IT system called "dealer management system," known as dealer management system (DMS). Generically, dealer management system (DMS) includes inventory tool kits to manage parts availability information while arranging appointments. Often, the same solution is applied to book available dates in the service diary, but inventory techniques are intended for basic control and future decisions are not supported properly.

Inventory models, particularly "newsvendor" solutions, built in the current dealer management systems (DMS) have been studied intensively. Unfortunately, inventory models have an important limitation as they are intended to obtain a point forecast [21], a mathematical expression to help in determining the economic order quantity [22], or the ordering frequency, to keep goods or services flowing to the customer without interruption [23] or delay [24]. Therefore, inventory models only deal with part of the problem of capacity reserve.

The second limitation of inventory models is, even when they can incorporate uncertain demand, the main application that is directed to quantify decisions or to estimate profit (or loss) of unsold units, and so forth, but, whatever the model is, there is a common goal: to maximize the expected profit [25]. An important constraint of inventory models is, according to the reviewed literature [26], the resource requirements that are not fully known when a decision about the service resource distribution is taken due to the nature of customer behaviour. Thus, a strategy that balances service quality and cost yields must be found [27].

A third limitation of inventory models is the dating process that frequently does not work properly given that customers are not always able to arrive within the appointed time window, delaying the reception operation and creating a bullwhip effect in upstream dates [28]. Therefore, inventory models cannot respond to the car manufacturer problem since they do not cover potential emergencies or capacity reserve. Additionally, in the passenger car industry, profit is not always related to stock trade but to a customer long-term relationship.

Yield management [29] is another methodology intended to manage the capacity of service systems. An important limitation is yield management that focuses on service pricing instead of service constraints and system capacity. This approach seems to be interesting for other service sectors, such as hotels and airlines, where the service duration is well known (i.e., one night, 3 hours flight) and service prices vary with the demand. Therefore, yield management models are not specifically developed to respond to the questions we aim to answer with our suggested service model for the passenger car industry.

In addition to the above methods, a common optimization methodology used in the service operation consists in running a simple forecast to estimate future demand values, without estimating uncertainty by means of a probability model [5]. Forecasts are then used to feed a mathematical expression that can be derived to minimize or maximize a variable. Nonetheless, there is no uncertainty quantification incorporated in the above optimization method as input demand is taken as an aggregated value, without differentiating between elective and nonelective demands.

In other cases [4], previous service's research mainly focuses on understanding and measuring customer expectations and perceptions about the quality of service being provided. This would result in ascertaining the gap between customers' expectation and perception. The obvious next stage is to identify the reasons for the gap between customer's expectation and service capacity and finally provide suggestions for bridging this gap and a follow-up of the effectiveness of the actions taken.

1.3. Research Purpose. In the context of demand uncertainty, resolution of optimal capacity is very strongly dependent on an appropriate specification of the service outputs. One limitation of previous studies is that they have used aggregate measures of service to define outputs [30]; a second limitation is the reliance on annual or quarterly fluctuations in demand to system responses to nonelective demand [31], but failing to take account nonelective demand leads to a misspecification of system cost output relation [32].

With the input desegregation in mind [33], just after the beginning of a time period, when the aggregated demand for this period is known, a decision can be made, but this works against optimal capacity. Thus, when capacity reserve is expensive or the rejection rate is high, any further increase of its value will cause a decrease in optimal capacity [34].

An important property of the time series is that constraints on elective and nonelective demand are separated from other constraints [35]. On the other hand, in hospital emergency room (ER) applications it is being assumed that all hospitals have similar patient stream structure and that patients arrive at the hospital according to a Poisson flow [36], but without taking into account how the stochastic nature of demand is related to the type of case being serviced, while in this paper we will incorporate this relationship. Thus, our research will implement queuing theory to study arrival patterns at the service reception, waiting lines and servers, waiting times, and tasks completed [37].

To summarize, this paper is about the stochastic simulation of the process of service capacity reserve in the passenger car industry. A stochastic model has been implemented in a Monte Carlo simulation code written in Matlab and has proved to be a very useful tool for optimal decision making under uncertainty, involving an optimization process to define and maximize new key process indicators (KPIs).

The major contributions of this paper are the definition of new key process indicators (KPIs) and the development of valid integrated capacity model to respond to the needs of the passenger car service industry.

2. Our Integrated Approach: Capacity Reserve Model

There is a very scarce literature on applying simulation techniques to capacity reserve in the car industry, and inventory models do not include customer expectations but are intended to define some constraints related to supply chain specifics. Our simulation process has been developed to fill this gap, finding the conditions for maximum average service occupancy and average minimum capacity reserve cost; the probability distributions are obtained from the Monte Carlo simulations.

According to specific studies [38], in practice, service operation algorithms are ultimately carried out by computer simulations. Therefore, the Markov chain usually simulated is only an approximation to the true chain. Such limitations affect the simulation process, reducing the final results and raising questions about the validity of the previous algorithms used to build the model.

In this paper, we propose a new methodology which substantiates the integration of existing strategies used in medical installations to develop a valid model to be used in the passenger car industry. The model will be used to predict unexpected service demand and set a decision support system to estimate (in accordance with the tradeoff of above) the optimal operational costs (service efficiency mentioned above) and optimum service capacity reserve (servicing unexpected customers) by coupling discrete events with simulation techniques.

Additional references [39–41] will be cited below when describing the methodology.

2.1. Basics of the Methodology. We propose a major innovation which implements changes in demand estimation for the initial inputs, providing definition of new outputs and the development of a stochastic simulator for the whole process.

Our methodology incorporates the risk in quantitative analysis and decision making; thus, we are able to provide service managers with a range of possible outcomes and the probability for each of them. Thus, we can select different simulated variables and compare with the logical solution of going for the most conservative decision; this is, keeping the service workshop layout as is, but considering the impact of increasing service staff.

The methodology follows the 5-stage process flow as displayed in Figure 1.

(i) *1st Stage: Service Demand Estimation*. We split the total demand (TV) in two major types, elective and nonelective, each with its own probabilistic distribution; therefore, we can integrate the stochastic models and sources of uncertainty of demand (elective and nonelective) and propagate this uncertainty to the output, thus adding value to predictions and allowing for statistical interpretation. We do this by applying ARIMA models to step ① in Figure 1, for both types, and specific models are developed for each. Input data is gathered from the dealer management system (DMS).

(ii) *2nd Stage: Service Times Definition*. Dealer's historic service data are gathered from the dealer management system (DMS) to estimate the probability distribution of each service time variable (reception, parts, and workshop). A stochastic queuing system is thus defined and fed with the adequate probabilistic models at step ② in Figure 1.

(iii) *3rd Stage: Service KPI Definition*. We define a new set of output key process indicators (KPIs) at step ③ in Figure 1, which are functions of the random variables defined in stages 1 and 2.

(iv) *4th Stage: Monte Carlo Simulation*. We run simulations to generate samples of the random variables of stages 1 and 2 and then propagate this uncertainty to obtain samples of the joint distribution of the different key process indicators (KPIs). If we do this for a number of scenarios (changing the number of technicians and of work bays), we will obtain different samples of the KPI joint distribution, one sample for each scenario. From each Monte Carlo sample we produce a report which displays a probabilistic analysis.

(v) *5th Stage: Results Analysis and Optimization*: Here we analyse the simulation results to identify the service operation conditions for maximum average service occupancy and average minimum capacity reserve cost; the different scenarios are given in terms of the number of work bays and technicians. During the optimization stage, our methodology incorporates other variables, that will be defined later, which are used to identify the optimal scenario to assess the current system's effectiveness and improve ability to anticipate the impact of various changes in the service settings, similarly to previous researches [13]. This will be discussed in Section 3.

2.2. Stage 1: Service Demand Variable Definition and Modelling. A common way to manage service times is to request the customer to arrange a valid date for the next visit in advance. We define elective demand (ED) as the total number of prearranged visits to the service. This variable increases service reception managing capacity and saves time and money by arranging parts and skilled staff in advance.

As discussed in the introduction, the dating process does not work properly if customers do not arrive in time or change their minds, delaying the reception operation and creating a bullwhip effect in upstream dates [28]. Some customers dislike the dating process due to the inflexibility and lack of same or next day availability. Elective demand (ED) is thus a stochastic variable subject to high variability, partially dependable on customer requests which are only known with certainty after the arrival. We define nonelective

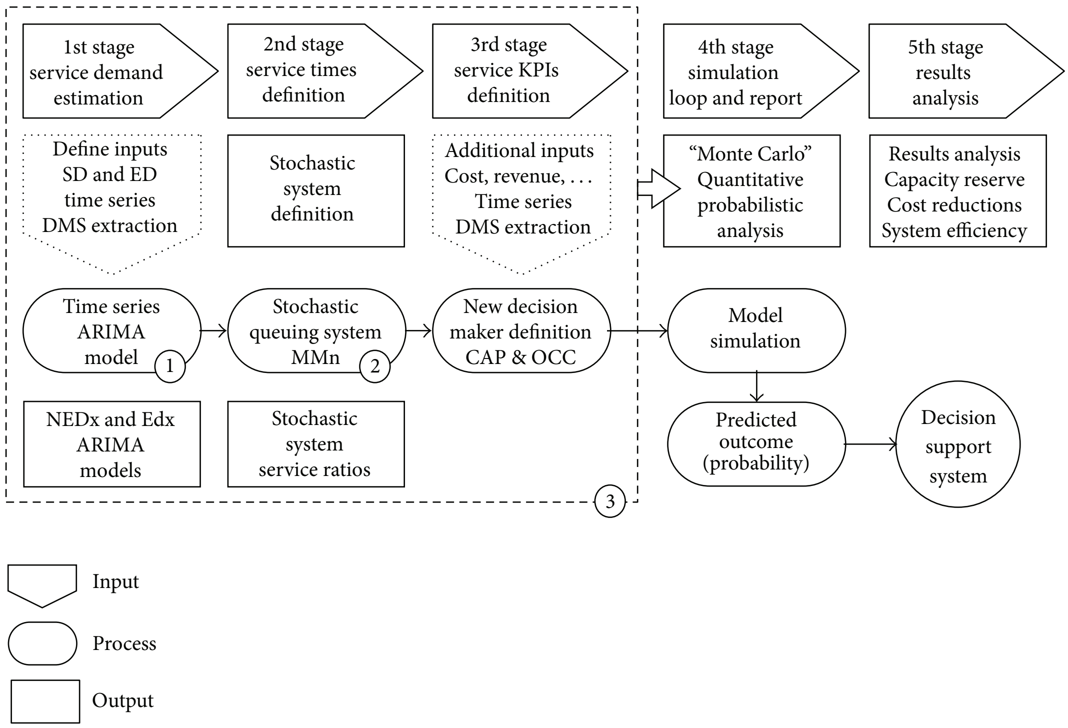

Figure 1: Methodology.

demand (NED) as the total number of unexpected services accepted in the system without a previous appointment.

As stated before, [26], resource requirements are not fully known at the time when a decision about the service resource distribution is taken. Therefore, in order to obtain valid forecasts, elective demand (ED) and nonelective demand (NED) should be the result of different random processes and can be expressed in a mathematical form as a probabilistic concept used to describe a sequence of random variables (stochastic process) that evolves in terms of another variable (index), usually time. Each of the random variables of the full service process has its own probability distribution, and variables may be correlated or not.

Hence, we suggest using ARIMA models for elective demand (ED) and nonelective demand (NED) forecasting. It should be noted there is a clear difference between both variables, and that is the reason to use separated variables (ED & NED) and feed the queuing system with each variable unique probabilistic distribution. This is required to demonstrate our methodology obtains valid results, integrating uncertainty and adding statistical value to the simple forecasting processes.

As discussed previously, nonelective demand (NED) has never been studied or estimated in the passenger car industry yet; therefore, there are no time series data to work with, neither the cost implicit in reserving capacity to service unexpected demand has been part of any research. This cost is a new concept which we will define below, in stage 3, as the capacity reserve cost to serve all nonelective demands (CAPRNED).

In the current economic situation, customers are demanding prompt and flexible service; thus, nonelective demand (NED) is becoming a huge issue for all passenger car brands. Elective demand (ED) and nonelective demand (NED) balance has a cost related to the capacity reserve to suit nonelective demand (NED) service needs.

If capacity reserve for nonelective demand (NED) is high, services could lose elective demand's (ED) service income and profit would be lower than expected. Also, if service is full with elective demand (ED) only, there is no capacity to suit nonelective demand (NED) arrivals and customers will turn away.

Else, operating at full capacity sets the optimal reserve capacity levels compatible with economically efficient utilization but imposes a cost, however, in the form of production inflexibility, leading to patients being queued or turned away. At the same time, failing to take account nonelective demand leads to a misspecification of system cost output relation [32].

As explained in the literature review, emphasis has been on estimating (and minimizing) the cost of maintaining reserve capacity rather than using nonelective demand as part of a decision support system. Our research will lean on the work referred to in this paragraph and extend it to apply the proposed methodology to unexpected demand at the car industry, according to the reviewed literature [12–18, 20].

Even when dealer management systems (DMS) are not focused on time series analysis, we can gather sufficient data to create time series to work with, which will include the following information:

(i) period,
(ii) total demand (TV),
(iii) elective demand (ED),
(iv) monthly work time,
(v) monthly invested time,
(vi) monthly invoiced time,
(vii) total part sales,
(viii) cost of part sales,
(ix) vehicle sales,
(x) cumulative vehicle car park.

Now, we can estimate the following parameters with the previous data:

(i) average invested hours per vehicle (invested time/car park),
(ii) average invoiced hours per vehicle (invoiced time/car park),
(iii) average sold parts per vehicle (total part sales time/car park),
(iv) average technician employment (invested time/work time),
(v) average technician productivity (invoiced time/work time),
(vi) average technician efficiency (invoiced time/invested time).

A simpler modelling could be developed by using an aggregated ARIMA model for total demand (TV) time series data only, but this would not allow us to relate the model to unexpected visits. Thus, in order to obtain a valid time series for nonelective demand we must gather the relevant data from the dealer management systems (DMS).

Dealer management systems (DMS) usually register all service visits but also manage the appointment process effectiveness by registering elective demand separately. We define total demand (TV) as the total number of visits in a given time period:

$$\mathrm{TV} = \sum (\mathrm{ED} + \mathrm{NED}) . \quad (1)$$

Total demand (TV) is thus the sum of elective and nonelective demands. Since this paper is intended to set a valid methodology to estimate the unexpected demand reserve costs, we need to differentiate between elective demand (ED) and nonnlective demand (NED) rather than using a single nondisaggregated variable, as total demand (TV) is.

Once a time series is available for total demand (TV) and elective demand (ED), we will be able to obtain the nonelective demand (NED) time series to predict future values with specific ARIMA models for each. This analysis will produce a forecast with uncertainty bands and confidence intervals that can be used to confirm if both time series forecasts are confident simultaneously.

The sample data time series from a real service is shown in Table 1; also data graph is displayed in Figure 2. Data will be used to estimate valid service demand ARIMA models, for elective demand (ED) and nonelective demand (NED), and to feed the stochastic queuing system.

An important property of the observed samples of time series is that constraints on elective demand (ED) and nonelective demand (NED) are independent of other constraints [35]. One axiom of our research is that elective demand (ED) and nonelective demand (NED) are independent. To check this, we must confirm from the data that elective demand (ED) and nonelective demand (NED) are not correlated.

Thus, individual ARIMA models and forecasts for elective demand (ED) and nonelective demand (NED) will have the form of a parametric expression that relates the future value to previous ones, plus the noise. Given an ARIMA model of consumer demand and the lead times at each stage, it has been proven that the orders and inventories at each stage are also ARIMA [28], and closed-form expressions for these models are given.

2.3. Stage 2: Service Times. We now deal with the second stage: we estimate the probability distributions required to feed a stochastic queuing system and emulate the whole service operation.

We build the queuing model to analyse the behaviour of the system along time and the reaction to different stimuli and waiting times for a queue in which customers require simultaneous service from a variable number of servers [41]. In previous studies, the service systems considered are centralized and controllable and do not generate labour at a constant rate [40].

Tasks are admitted upon generation and processed by the system and completed labour is ejected from the system that has the capability of dealing with as many jobs per unit time on average as possible. Under this generic framework the system capability is measured as the maximum rate of work arrivals for which the system has a steady state [39].

We differ from the previous statement since we are considering service operation as a complete unit; that is, we include additional departments and not just service's workshop. This is, we are considering Parts and Reception times, including reception delays due to customer unavailability to arrive, "elective" customers changing to "nonelective," and other delays related to parts ordering and delivery.

We propose to measure the full service system capability (as shown in Figure 3) by running a queuing model built in the simulation loop at a constant arrival rate of work (arrival rate λ: shown in Figure 3 and defined in Table 2). Therefore, our methodology will cover a unique service cycle (as seen by customers) with the following phases, shown in Figures 2 and 3:

(1) arrival,
(2) reception,
(3) parts,
(4) service workshop.

Table 1: Sample data.

Period	TV	ED	NED
1	107	91	16
2	94	91	3
3	131	119	12
4	116	113	3
5	172	172	0
6	184	182	2
7	154	148	6
8	75	75	0
9	107	97	10
10	125	107	18
11	129	108	21
12	103	91	12
13	149	135	14
14	130	130	0
15	165	147	18
16	249	248	1
17	137	137	0
18	179	179	0
19	145	145	0
20	117	117	0
21	195	179	16
22	149	131	18
23	205	169	36
24	135	122	13
25	150	121	29
26	149	115	34
27	144	140	4
28	132	127	5
29	252	243	9
30	165	159	6
31	169	144	25
32	105	94	11
33	172	142	30
34	131	109	22
35	180	144	36
36	128	115	13
37	180	141	39
38	148	127	21
39	155	127	28
40	138	124	14
41	258	222	36
42	168	152	16
43	151	135	16
44	114	99	15
45	128	97	31
46	151	140	11
47	139	111	28
48	97	85	12
49	160	136	24
50	173	149	24

Table 1: Continued.

Period	TV	ED	NED
51	146	119	27
52	331	307	24
53	140	129	11
54	137	124	13
55	174	163	11
56	88	86	2
57	131	117	14
58	183	162	21
59	135	120	15
60	125	119	6
61	98	93	5
62	145	124	21
63	138	116	22
64	200	185	15
65	226	212	14
66	166	156	10
67	151	139	12
68	124	99	25
69	156	137	19
70	183	165	18
71	141	118	23
72	118	109	9
73	119	111	8
74	122	109	13
75	113	95	18
76	221	197	24
77	193	171	22
78	140	133	7
79	160	144	16
80	117	92	25
81	141	130	9
82	148	139	9
83	135	123	12
84	111	99	12
85	108	105	3
86	110	109	1
87	112	106	6
88	132	131	1
89	128	126	2
90	129	127	2
91	117	109	8
92	95	91	4
93	117	103	14
94	119	116	3
95	134	126	8
96	136	134	2
97	112	111	1
98	70	61	9
99	89	82	7
100	84	81	3

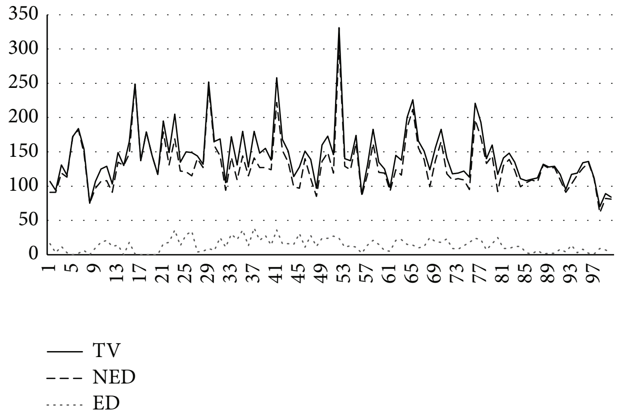

Figure 2: Sample data plot.

The arrival rate (λ) is given by the elective demand (ED) and nonelective demand (NED) forecasts.

Now the purpose of this section is to do the following.

(1) We use the dealer management system (DMS) data to estimate the probability distributions of the following random variables:

(a) reception time,
(b) parts time,
(c) workshop time.

(2) Subsequently, we use these distributions to feed our queuing system and estimate the distribution of its random variables, in terms of the reception time, parts time, and workshop time described referred in the previous point:

(a) customer time at the reception of the service,
(b) parts lead time,
(c) parts availability,
(d) working time per vehicle,
(e) additional specific queuing model parameters (defined below in Section 2.3.1 Queuing Methodology for Service Times).

We define the total preparation time (TSR) as the necessary time in minutes to deal with the customer, find the required parts, and take the vehicle to the technician. This variable is not measured in the industry and requires physical checks on the field and data sampling to understand its structure. The total preparation time (TSR) is thus the result of adding the time the customer is at the dealership and the time to get the parts physically:

$$\text{TSR} = \text{TCust} + \text{TParts}, \tag{2}$$

where

(i) TCust (customer waiting time) is the time to manage customer request at the reception desk and raise a job card;

(ii) TParts (parts lead time) is the time to get the parts supplied before being fitted to the car in the service.

Now, knowing the car fleet for a particular region we estimate from the dealer management system (DMS) data the distribution of the service time per vehicle (TSW), which will vary with the model, region, service skills and competency, workshop layout, and other parameters.

2.3.1. Queuing Methodology for Service Times. As we discussed in the introduction, queuing theory allows for the study of waiting lines and servers, including arriving patterns at the queue, waiting times, and tasks completed [37].

We build a queuing system into the Matlab code, with the following 6 characteristics.

(1) Arrival pattern of customers: as mentioned before, it is a constant rate process, where the rate is a function of the total demand (TV) and therefore will depend on the elective demand (ED) and nonelective demand (NED) probability distributions.

(2) Service pattern of customers: it depends on the number of customers queuing for service and will be a function of the distribution of the customer waiting time (TCust).

(3) Queue discipline: in our research we set priorities in terms of part availability. If a part is backordered, the customer will be requested to wait. Therefore, this will be a function of the distribution of the parts lead time (TParts).

(4) Queuing capacity: it is limited by the number of appointments plus the emergency visits. It depends on the customer waiting time (TCust) and the parts lead time (TParts). Thus, it is a function of the distribution of the total preparation time (TSR).

(5) Number of servers: rather than considering a two-stage server system (reception and workshop), we set our system as a single-level server, where customers leave their vehicles at the reception but they do not physically wait until it is taken to the workshop. Therefore, this characteristic will be a function of the service time per vehicle (TSW). Maximum number of servers is given by the facility layout and could vary with time depending on the technician's availability, including holiday periods, sick leaves, and training courses. We will simulate this variability as part of our methodology.

(6) Number of work phases along the complete service process: similar to hospital's emergency room (ER) we assume a single stage service for the whole service process, but we simulate the time variability due to the work complexity and different kinds of services Thus, it is a function of the service time per vehicle (TSW).

In Figure 4, customers ($C1, C2, \ldots, Cc$) arrive at the reception area; they could be part of elective demand (ED) or nonelective demand (NED) with their own probability

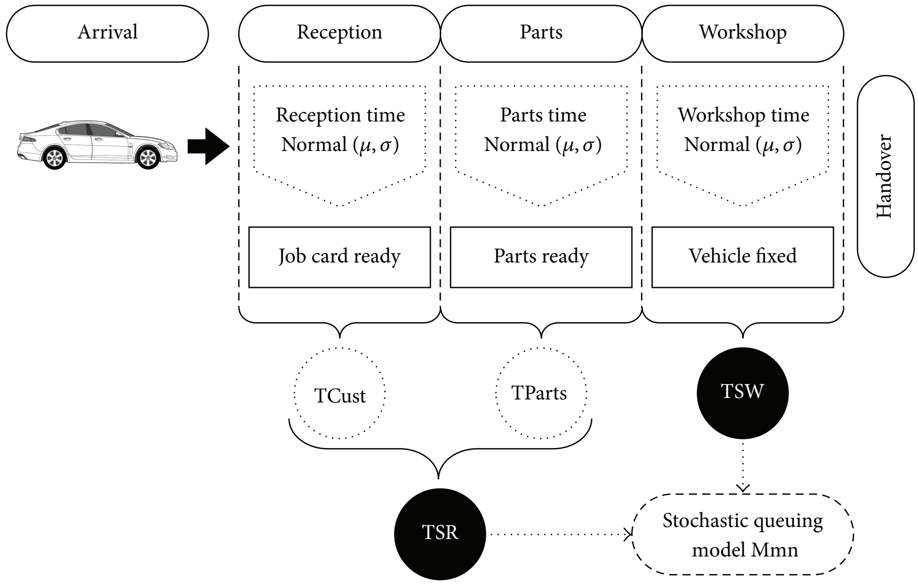

Figure 3: Service process times: total preparation time (TSR) and service time per vehicle (TSW).

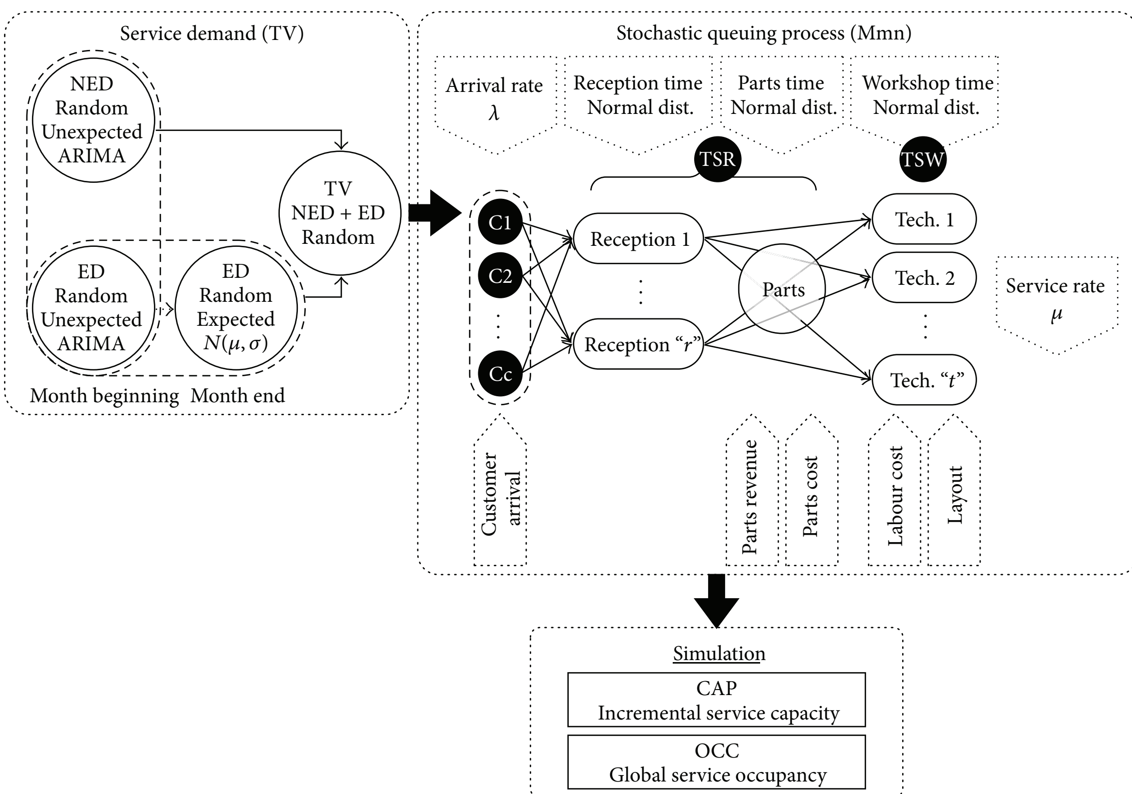

Figure 4: Theoretical process for capacity reserve and service occupancy estimation.

distribution, but once at the service reception, they travel through the network and are served at the reception nodes (receptions $1, 2, \ldots, r$) through parts and to the workshop nodes (Tech. $1, 2, \ldots, t$).

The first variable to be defined is the customer's arrival rate (λ), where Cc is the number of customers waiting at the queue, where, as an open network, customers can join and leave the system as shown in Figure 4.

Arrival rate (λ) and service rate (μ) are function random variables: the total preparation time (TSR) and the service time per vehicle (TSW), respectively; therefore, they are random variables.

Table 2: Queuing system steady-state measures of effectiveness.

Steady-state measures of effectiveness	Variable	Name
$\text{Nveh} = \dfrac{\text{TV}}{\text{WDM}}$	Number of vehicles arriving to the service	Nveh
Simulation variable	Average number of vehicles in the Q-NTec	L
Simulation variable	Average number of vehicles in the Queue	Lq
$\text{FTI} = 1 - \dfrac{\text{Wtecx}}{\text{NTec}} = P_0 + \left(\dfrac{s-1}{s}\right) \times P_1 + \cdots + \left(\dfrac{1}{s}\right) \times P_{S-1}$	Fraction of time that a technician is idle	FTI
$\text{VhW} = \text{NVeh} - L$	Average number of vehicles that are being worked	VhW
$\text{Wtec} = L - \text{Lq} = \text{NTec} - \sum_{n=0}^{s-1} (s-n) * P_n$	Average number of technicians that are working	Wtec
$\text{EFC}_{\text{Tec}} = \dfrac{\text{VhW}}{\text{NVeh}}$	Operating efficiency per vehicle	EFC_{Veh}
$\text{EFC}_{\text{Veh}} = \dfrac{\text{Wtec}}{\text{NVeh}}$	Operating efficiency per technician	EFC_{Tec}
$\mu = \dfrac{1}{\text{TSW} \times 60}$	Service rate [vehicles/hour]	μ
$\lambda = \dfrac{1}{\text{TSR} \times 60}$	Arrival rate [vehicles/hour]	λ
$P_n = \dfrac{\text{NVeh}!}{h! \times (\text{NVeh} - h)!} \times \left(\dfrac{\lambda}{\mu}\right)^h \times P_0$	Probability n vehicles are in the queuing system (Q-Ntec)	Pn
$P_n = \dfrac{1}{(P_0)^{-1}}$	Probability of no calling units in the queuing system.	Po
Initial Po_inverse = 0, then: $P_0 = (P_0)^{-1} + \text{den\_1}$	Po inverse	Po^{-1}
$\text{f(den\_1)} = \dfrac{\text{NVeh}!}{h! \times (\text{NVeh} - h)!} \times \left(\dfrac{\lambda}{\mu}\right)^h$	Operator	f(den_1)

In queuing theory, the state of the system is given by a vector with different variables. The complete list of variables to be used in the stochastic queuing system (defined by the above characteristics) is detailed in Table 2.

The queuing system allows us to estimate the average cycle service time per vehicle (AT_{PV}), this time is not just calculated adding the service time per vehicle (TSW) and the total preparation time (TSR). It is the result of a complex forecast process to estimate the whole service time, including timing delays due to operational inefficiencies and other system limitations; that is, vehicle movements included in the Service Time per Vehicle (TSW) or some Elective Demand (ED) missing the time window appointment.

The time spent in each of the processes above is a random variable, with its own probability distributions and parameters.

2.4. Stage 3: Service Key Process Indicators (KPIs). This is an essential contribution of this research and consists of the definition of a set of new KPIs for the passenger car industry, in terms of the inputs defined above and of additional random variables and parameters to be defined as follows. A comprehensive list of key process indicators (KPIs) is detailed in Table 3, showing input variables to feed the model with and output variables to be obtained from the Monte Carlo simulation.

Traditionally, service management organizes and measures the technician's time to administer the service department's labour availability and performance to maximize operational net profit. Technician performance and time control are basically monitored upon the 3 measurement ratios below.

(i) Productivity: time the technician is physically present at work divided by the actual working hours.

(ii) Efficiency: time spent working on a vehicle divided by the flat rate time received.

(iii) Availability: flat rate time received divided by the time the technician is physically present.

Without doubt, these three indicators are useful to monitor technical service performance, but they do not incorporate delays on reception or parts backorder. Therefore, we need to define new measures of effectiveness which take into account inputs from service, parts, and reception and include customer "expectations" in our model. We now focus on the

TABLE 3: List of acronyms: parameters and variables-simulation inputs and outputs.

(a)

Acronym	Definition	Source	Parameter
Lab	Retail labour rate	DMS data	Constant
Ntec	Number of available technicians		Constant
WBN	Total number of staffed work bays		Constant
Ψ	Number of work bays per technician (WBN/Ntec)		Constant (from 1 to 2)

(b) Simulation Random Inputs

Acronym	Definition	Source	Prob. distribution
ED	Elective demand	DMS data	ARIMA
HPRES	Working time per month (hours)		$N(399{,}65; 85{,}92)$
InvoT	Invoiced time		$N(373{,}76; 76{,}9)$
NED	Nonelective demand (NED = TV − ED)		ARIMA
Nveh	Number of vehicles arriving to the service per day		$N(6{,}13; 1{,}85)$
PartsCost	Cost of parts sale		$N(21297{,}95; 11963{,}79)$
PS	Parts sale		$N(27278{,}41; 13517{,}82)$
TV	Total demand		ARIMA
WDM	Working days per month		$N(21; 1)$
WTD	Daily working time (hours per day)		$N(6{,}51; 1{,}13)$

(c) Simulation random outputs

Acronym	Definition	Source
ATPV	Average cycle service time per vehicle (hours)	Queuing + simulation
CAP	Service incremental capacity	
CAPRNED	Capacity reserve cost to serve all nonelective demands	
CAPRNED_1	Capacity reserve cost to serve 1 unexpected vehicle	
CEW	Empty work bay cost estimation	
DMS	Dealer management system	
EFC	Service system efficiency	
EFC_{Tec}	Operating efficiency per technician	
EFC_{Veh}	Operating efficiency per vehicle	
GSR	Monthly gross service revenue estimation	
InvoTn	Invoiced time estimation	
OCC	Service occupancy	
PartsCostn	Cost of parts sale estimation	
PFunit	Profit per vehicle in service	
PSn	Parts sale estimation	
Pstec	Parts sale per technician	
Tcust	Customer waiting time	
TEXP	Monthly workshop total cost estimation	
Tparts	Parts lead time	
TPF	Service total gross profit estimation	
TSR	Total preparation time. TSR = TCust + TParts	
TSW	Service time per vehicle	
VhW	Average number of vehicles that are being worked	
Wtec	Average number of technicians that are working	

(d) Other queuing system outputs

Acronym	Definition	Source
f(den_1)	Mathematical operator	Queuing + simulation
FTI	Fraction of time that a technician is idle	
L	Average number of vehicles in the Q-NTec	

(d) Continued.

Acronym	Definition	Source
Lq	Average number of vehicles in the queue	
P_n	Probability n vehicles are in the queuing system (Q-Ntec)	
Po	Probability of no calling units in the queuing system	
Po^{-1}	Po inverse	
λ	Arrival rate [vehicles/hour]	
μ	Service rate [vehicles/hour]	

seven new key process indicators (KPIs) that we are going to define:

(i) service incremental capacity (CAP),

(ii) service occupancy (OCC),

(iii) nonelective demand (NED),

(iv) cost of capacity reserve (RCAP),

(v) cost of empty work bay (CEW),

(vi) capacity reserve cost to serve all nonelective demand (CAPRNED),

(vii) profit per vehicle in service ($\mathrm{PF}_{\mathrm{UNIT}}$).

2.4.1. Service Incremental Capacity. We define the service incremental capacity (CAP) as the system potentiality to admit additional workload without interrupting on-going works. In other words, the service incremental capacity (CAP) is the capability to accommodate unexpected customers during the normal working time.

Service incremental capacity (CAP) is calculated as a percentage rate of the whole service process capacity that will decrease as long as the number of vehicles through the system increases:

$$\mathrm{CAP} = \frac{\mathrm{WBN} \times \mathrm{WTD}}{\mathrm{ATPV}} \times \Psi \times \mathrm{EFC}. \quad (3)$$

Therefore, we express the service incremental capacity (CAP) as a function of the following.

(i) WBN is the total number of staffed work bays. It is constant and depends on the workshop layout.

(ii) WTD is the daily working time.

(iii) ATPV is the average cycle service time per vehicle.

(iv) Ψ is the number of work bays/technician, which usually can vary from 1 to 2. It is constant and will depend on the facility layout.

(v) EFC is the average workshop efficiency rate.

2.4.2. Service Occupancy. We define the service occupancy (OCC) as the measurement of vehicles in-progress through the service system. It is related to the number of vehicles entering the system divided by the work bays and the system capability expressed as service incremental capacity (CAP).

Service occupancy (OCC) is calculated as a percentage rate of the whole service process workload that will grow as long as the number of vehicles through the system increases.

$$\mathrm{OCC} = \frac{\mathrm{TV}}{\mathrm{CAP} \times \mathrm{WDM}}. \quad (4)$$

OCC is the service occupancy expressed above as a function of the following.

(i) Total demand (TV) is the total visit number in a given time period.

(ii) WDM is the total working days in a given month.

Now, if we write nonelective demand (NED) as a function of (1) and (4), then

$$\mathrm{NED} = (\mathrm{OCC} \times \mathrm{CAP} \times \mathrm{WDM}) - \mathrm{ED}. \quad (5)$$

2.4.3. Cost of Capacity Reserve. We define cost of capacity reserve (RCAP) as the opportunity cost in € to reserve service capacity in the form of empty work bay. It is expressed as the following equation:

$$\mathrm{RCAP} = \frac{\mathrm{TPF}}{\mathrm{TV}} + \frac{\mathrm{TEXP}}{\mathrm{TV}}, \quad (6)$$

where

(i) TPF is the monthly workshop gross profit,

(ii) TEXP is the monthly workshop total cost.

The first component of the above expression is the profit per unit ($\mathrm{PF}_{\mathrm{UNIT}}$):

$$\mathrm{PF}_{\mathrm{UNIT}} = \frac{\mathrm{TPF}}{\mathrm{TV}}. \quad (7)$$

By replacing (7) with (6), we define the capacity reserve cost for a single nonelective demand ($\mathrm{CAPRNED}_1$):

$$\mathrm{CAPRNED} = \mathrm{PF}_{\mathrm{UNIT}} + \frac{\mathrm{TEXP}}{\mathrm{NED}}. \quad (8)$$

Also, by replacing (5) with (8),

$$\mathrm{CAPRNED}_1 = \mathrm{PF}_{\mathrm{UNIT}} + \frac{\mathrm{TEXP}}{(\mathrm{OCC} \times \mathrm{CAP} \times \mathrm{WDM}) - \mathrm{ED}}. \quad (9)$$

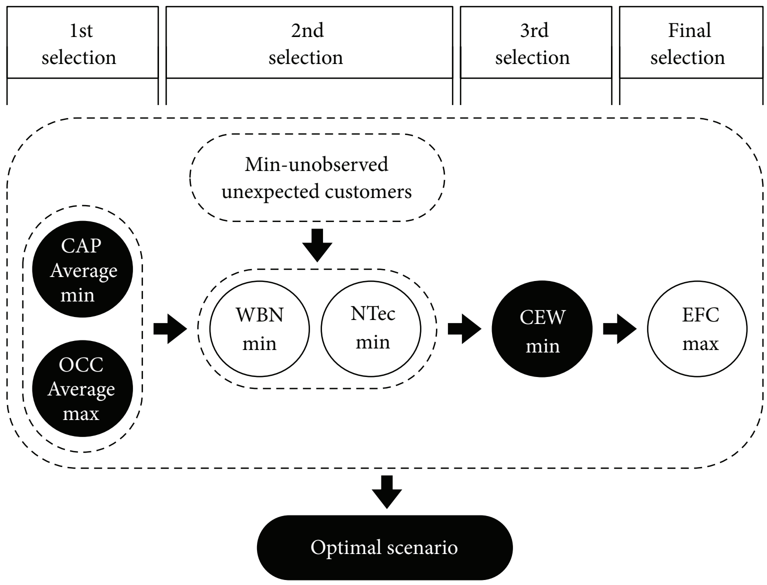

Figure 5: Optimal scenario selection criteria in 5 levels.

Therefore, the capacity reserve cost to serve all nonelective demand (CAPRNED) will be

$$\mathrm{CAPRNED} = \mathrm{NED} \times \mathrm{CAPRNED}_1. \tag{10}$$

And then replacing (8) with (10),

$$\mathrm{CAPRNED} = \left(\mathrm{NED} \times \mathrm{PF}_{\mathrm{UNIT}}\right) + \mathrm{TEXP}. \tag{11}$$

Given that service incremental capacity (CAP), service occupancy (OCC), nonelective demand (NED), and capacity reserve cost to serve all nonelective demand (CAPRNED) are function of the random variables detailed in Table 3, they are also random variables whose joint distribution will be estimated through Monte Carlo simulation.

2.5. Stage 4: Monte Carlo Simulation. The purpose of this section is to obtain Monte Carlo samples of the distributions of the key process indicators (KPIs) defined above.

As covered in the literature review, the queuing system, or Markov chain, actually simulated is only an approximation of the true chain. Such results with finite precision and range are introduced and pose further questions about the validity of these algorithms [38]. Thus, we apply the Monte Carlo methodology to generate samples of the input variables defined in Section 2 and propagate their uncertainty obtaining samples of the distributions of the key process indicators (KPIs).

This stage has been implemented in a Matlab code which runs a complete set of calculations to simulate capacity reserve with a queuing model to obtain samples of the same size for each KPI.

We repeat the simulations for a number of scenarios (changing number of technicians and of work-bays). The Matlab code will run two loops taking different values of both variables (number of technicians and of number of work-bays). The value of each variable will be increased 1 by 1 each loop to obtain different samples of the joint distribution of the key process indicators (KPIs). The sequence is as follows:

number of technicians → number of work bays → queuing system loop.

For each KPI, the simulation will then store the results in a tridimensional matrix: the first index varies with the sample (i.e., from 1 to 1000), the second with the number of technicians, and the third with the number of work bays.

3. Statistical Analysis of Key Process Indicators and Optimization

The purpose of this section is twofold:

(a) statistical analysis of the Monte Carlo samples of the joint distributions of our new key process indicators (KPIs);

(b) set the optimization criteria to define the optimum scenario in terms of recommended number of technicians and number of work bays.

After running the simulation we can estimate the joint and marginal distributions of the key process indicators (KPIs) and identify if there are additional relationships among them. A further statistical analysis is discussed in Section 4.

After the previous stages have been fully completed, our Matlab code identifies the service optimal scenario as a tradeoff between the dealer's total demand (TV) and the capacity reserve cost to serve all nonelective demands (CAPRNED).

Table 4: Simulation-decision support system final report.

(a)

Description	Constant	
Working days per month	WDM	21
Nonelective demand	NED_1	12
Work bays per technician	$WBtec_2$	1,25

(b)

Description	Variable	Average	Standard deviation
Customer waiting time (min)	TCust	37,61	12,966
Pats lead time (min)	TParts	3586,87	2080,052
Total preparation time (min)	TSR	3624,48	2080,052
Service time per vehicle (min)	TSW	6533,74	3730,76
% Time a technician is idle	FTI	16,89%	0,24
Working time per day (hours)	WTD	7,5	0,58
Parts sold per technician (€)	PStec	4755,03	1458,17
% Time sold at retail price	RLab	72,40%	0,072
Effective labour price (€)	ELab	66,13	0,486

(c)

Description	Simulation results (Ntec)			Simulation results (Ntec + 1)		
	Variable	Average	Standard deviation	Variable	Average	Standard deviation
Efficiency per vehicle	$EFCvh_1$	34,60%	0,25	$EFCvh_2$	36,31%	0,24
Average service efficiency	EFC_1	86,50%	0,23	EFC_2	83,11%	0,24
Workshop available time (min)	$HPRES_1$	1102,43	85,3	$HPRES_2$	1259,92	97,48
Average cycle service time per vehicle (hours)	$ATPV_1$	3,46	0,95	$ATPV_2$	3,79	1,13
Service incremental capacity	CAP_1	26,80%	—	CAP_2	20,50%	—
Service occupancy	OCC_1	49,00%	—	OCC_2	64,00%	—
Profit per vehicle (€)	$PFunit_1$	120,84	36,98	$PFunit_2$	137,83	42,27
Estimation of parts sale (€)	PS_1	33285,2	10207,19	PS_2	38040,2	11665,36
Work bays per technician	$WBtec_1$	1,43	0	$WBtec_2$	6,65	1,917
Cost of empty bay (€)	CEW_1	34,27	6,301	CEW_2	39,07	7,191
Capacity reserve cost for all nonelective demands (€)	$CAPRNED_1$	1861,32	449,97	$CAPRNED_2$	2122,85	514,798
Capacity reserve cost to serve 1 unexpected customer (€)	$CAPRNED_{11}$	155,11	37,498	$CAPRNED_{21}$	176,9	42,9
Number of cars in the workshop	VhW_1	4,5	3,206	VhW_2	4,72	3,15

The optimal scenario will be selected following the hierarchical approach displayed in Figure 5 and detailed below.

(1) The code will select those scenarios with a maximum service occupancy (OCC) and minimum capacity reserve cost to serve all nonelective demands (CAPRNED).

(2) Then it will order the selected scenarios starting from

   (a) the lowest number of work bays (layout constraints and cost),

   (b) the lowest number of technicians (operational constraints and cost),

   (c) the lowest number of unobserved unexpected customers.

(3) A 3rd level will filter those scenarios with the lowest cost for an empty server (work bay).

(4) The final level filters and selects the scenario with the highest whole service (reception, parts, and workshop) operational efficiency out of the previous selection. With the information from the "optimal

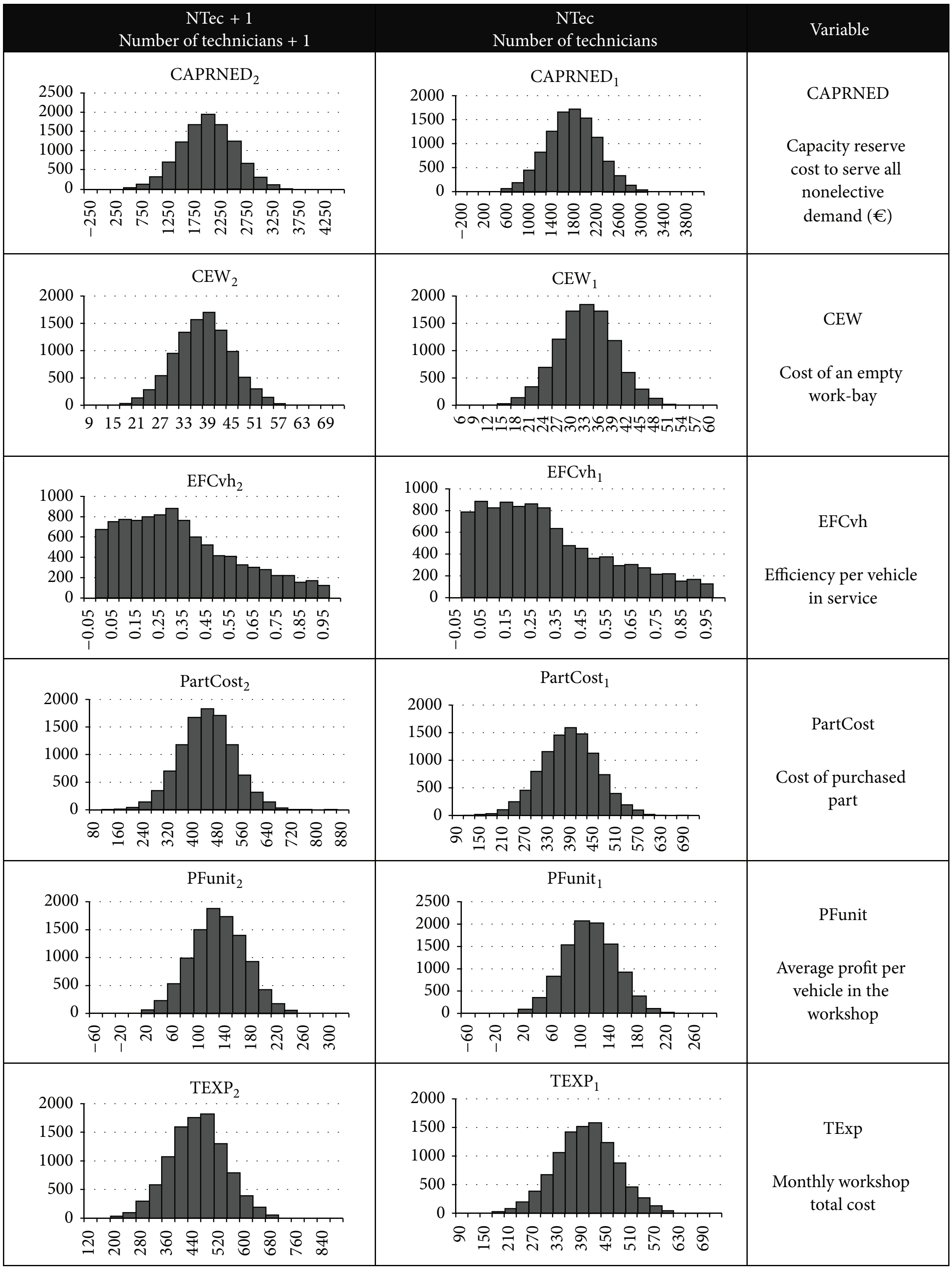

Figure 6: Simulation-output variables histograms.

Table 5: 95% confidence intervals.

(a)

Description	Input variable	Mean	Standard error	95% Confidence intervals	
				Lower limit	Upper limit
Elective demand	ED	147,859	4,32304	139,305	156,414
Nonelective demand	NED	13,3672	0,824012	11,7366	14,9978
Total demand	TV	161,328	4,51834	152,286	170,168

(b)

Description	Input variable	Sigma	Lower limit	Upper limit
Elective demand	ED	48,9096	43,5628	55,7652
Nonelective demand	NED	9,32263	8,30348	10,6294
Total demand	TV	51,1192	45,5308	58,2845

(c)

Output variable	Output variable	Mean	Standard error	Lower limit	Upper limit
Capacity reserve cost for all nonelective demands (€)	$CAPRNED_1$	1861,14	4,5	1852,31	1869,96
	$CAPRNED_2$	2122,64	5,15	2112,54	2132,74
Cost of an empty work bay (€)	CEW_1	34,26	0,06	34,14	34,39
	CEW_2	39,07	0,07	38,93	39,21
Efficiency per vehicle in service (%)	$EFCvh_1$	0,35	0,001	0,34	0,35
	$EFCvh_2$	0,36	0,001	0,36	0,37
Cost of purchased parts (€)	$PartCost_1$	402,27	0,76	400,79	403,75
	$PartCost_2$	458,65	0,86	456,96	460,34
Average profit per vehicle in the workshop (€)	$PFunit_1$	120,83	0,37	120,11	121,56
	$PFunit_2$	137,82	0,42	136,99	138,65
Monthly workshop total cost (€)	$TExp_1$	411,18	0,76	409,69	412,66
	$TExp_2$	468,83	0,86	467,14	470,52

(d)

Output variable	Output variable	Sigma	Lower limit	Upper limit
Capacity reserve cost for all nonelective demands (€)	$CAPRNED_1$	450,33	444,18	456,66
	$CAPRNED_2$	515,21	508,17	522,45
Cost of an empty work bay (€)	CEW_1	6,31	6,22	6,4
	CEW_2	7,2	7,1	7,3
Efficiency per vehicle in service(%)	$EFCvh_1$	0,25	0,24	0,25
	$EFCvh_2$	0,24	0,24	0,25
Cost of purchased parts (€)	$PartCost_1$	75,54	74,51	76,6
	$PartCost_2$	86,22	85,04	87,43
Average profit per vehicle in the workshop(€)	$PFunit_1$	37	36,49	37,52
	$PFunit_2$	42,29	41,71	42,88
Monthly workshop total cost (€)	$TExp_1$	75,71	74,68	76,78
	$TExp_2$	86,42	85,24	87,64

scenario," we will produce a final report (Table 4) which displays the expected key process indicators (KPIs) for the recommended number of work bays (WBN) and the recommended number of technicians (NTec). It also displays a second simulation with an additional technician (NTec + 1) to support the decision making in the following scenarios:

(i) Short term: identifying how indirect operational revenue or cost can be improved by increasing the operational staff.

(ii) Medium term: assessing the effectiveness of the current service system and identifying the impact of applying changes to the original service settings.

The report identifies also capacity and occupancy levels for the optimal scenario and how they could be affected when the existing staffs are increased by 1 head, provided that there is at least 1 additional work bay to be used for servicing nonelective demand (NED).

Also, we noticed when service staff is increased and unexpected demand is part of the capacity reserve, the empty work bay cost estimation is also reduced. As stated before, service incremental capacity (CAP) is inversely proportional to the service productive headcount, as it drops as soon as staff increases, while service occupancy (OCC) is directly proportional to the service staff.

This makes sense and confirms the expected outcome; the potential capacity we could have in the system should be lower when an additional vehicle is processed in the service system, showing an increment in the system occupancy rate. The simulation results suggest that services generate costs when reserving service capacity to serve nonelective demand.

4. Simulation Results Analysis and Uncertainty

This section shows the simulation results report in Table 4, including the relevant key process indicators (KPIs) information, and a sample of variables histograms in Figure 6. As said before, we use probabilistic distributions to demonstrate that our methodology obtains valid results integrating uncertainty and adding value to the simple forecasting processes which are common in the passenger car service industry.

In addition to the standard deviations shown below, interquartile ranges or 95% intervals could be easily computed from the samples as complementary measures of uncertainty. This is done in Table 5, where we are showing 95% confidence intervals for the means and standard deviations of each of the selected variables.

Other key process indicators (KPIs) are used to understand how economic variables could change depending on the solution applied, comparing the current layout and situation with the possibility of increasing service staff in one head.

Thus, the cost of empty work bay (CEW_1 and CEW_2 in Table 5) is increased as long as the profit per unit (PFunit_1 and PFunit_2 in Table 4) raises, so the cost implicit in capacity reserve to suit customer needs affects all the economic factors as we wanted to demonstrate.

Now we will study six variables out of the total number displayed in Table 4 (see Figure 6 and Table 5).

5. Conclusions

This paper studies a new approach, where, by analysing nonelective demand (NED), the apparent inefficiency resulting from services operating within production limits is understood. This analysis could also help brand managers when setting efficiency objectives with adequate adjustment for unexpected demand and its impact on cost structures.

We confirm here how separating nonelective demand (NED) from elective demand (ED) when estimating service costs is of paramount importance, as well as for labour fees setting and service level, which will depend also on how accurate service demand and general costs predictions are. Furthermore, the leftmost column of the simulation report, as displayed in Table 4, identifies some apparent inefficiencies resulting from services operating within production limits. With this information, the report compares several service process indicators to demonstrate how results can be affected by hiring additional technical staff.

The significance of the contribution of our research is the definition of new key process indicators (KPIs) to be used as a management tool for services. The capacity reserve strategy has been proved to be plausible and consistent, according to the reviewed literature of the hospital's emergency room (ER) field, with our conceptual arguments relating to production responses to demand uncertainty. Therefore, the information used allows for a more detailed specification of service output that can be applied to the passenger car industry to forecast service requisites and plan brand strategies which are aligned with the customer's real demand.

Future research could afford an exhaustive analysis to the data gathered after the Monte Carlo simulation. This could be done with the support of any of the existing statistical software packages to fully understand the existing relations among the multiple key process indicators (KPIs) and between inputs and outputs, like partial correlations and stochastic dependence between the new key process indicators (KPIs) defined in this paper.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] M. A. Cohen, N. Agrawal, and V. Agrawal, "Winning in the after market," *Harvard Business Review*, vol. 84, no. 5, pp. 129–138, 2006.

[2] R. N. Bolton, "A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction," *Marketing Science*, vol. 17, no. 1, pp. 45–65, 1998.

[3] V. Srivastava and A. Tyagi, "The study of impact of after sales services of passenger cars on customer retention," *International Journal of Current Research and Review*, vol. 5, no. 1, pp. 127–132, 2013.

[4] R. Katarne, S. Sharma, and J. Negi, "Measurement of service quality of an Automobile Service Centre," pp. 286–291, Proceedings of the 2010 International Conference on Industrial Engineering and Operations Management, Dhaka, Bangladesh, January 2010.

[5] M. A. Cohen, N. Agrawal, and V. Agrawal, "Achieving breakthrough service delivery through dynamic asset deployment strategies," *Interfaces*, vol. 36, no. 3, pp. 259–271, 2006.

[6] S. S. AmbekarIndian, "Service quality gap analysis of automobile service centers," *Indian Journal of Research in Management, Business and Social Sciences*, vol. 1, no. 1, pp. 38–41, 2013.

[7] A. Berndt, "Investigating service quality dimensions in South African motor vehicle servicing," *African Journal of Marketing Management*, vol. 1, no. 1, pp. 1–9, 2009.

[8] M. Hanif, S. Hafeez, and A. Riaz, "Factors affecting customer satisfaction," *International Research Journal of Finance and Economics*, vol. 63, no. 60, pp. 44–52, 2010.

[9] A. Herrmann, F. Huber, and A. Gustafsson, "From value orientated quality improvement to customer satisfaction: a

case study for passenger cars," in *Customer Retention in the Automotive Industry*, pp. 93–115, Gabler, 1997.

[10] A. A. Jahanshahi, M. A. Hajizadeh Gashti, S. A. Mirdamadi, K. Nawaser, and S. M. Sadeq Khaksar, "Study the effects of customer service and product quality on customer satisfaction and loyalty," *International Journal of Humanities and Social Science*, vol. 1, no. 7, pp. 253–260, 2011.

[11] M. D. Johnson, A. Herrmann, F. Huber, and A. Gustafsson, "An introduction to quality, satisfaction, and retention—implications for the automotive industry," in *Customer Retention in the Automotive Industry*, pp. 1–17, Gabler, 1997.

[12] R. D. Kamenetzky, L. J. Shuman, and H. Wolfe, "Estimating need and demand for prehospital care," *Operations Research*, vol. 30, no. 6, pp. 1148–1167, 1982.

[13] M. A. Badri and J. Hollingsworth, "A simulation model for scheduling in the emergency room," *International Journal of Operations and Production Management*, vol. 13, no. 3, article 13, 1993.

[14] Y. Gerchak, D. Gupta, and M. Henig, "Reservation planning for elective surgery under uncertain demand for emergency surgery," *Management Science*, vol. 42, no. 3, pp. 321–334, 1996.

[15] M. Bazargan, S. Bazargan, and R. S. Baker, "Emergency department utilization, hospital admissions, and physician visits among elderly African American persons," *Gerontologist*, vol. 38, no. 1, pp. 25–36, 1998.

[16] S. C. Brailsford, V. A. Lattimer, P. Tarnaras, and J. C. Turnbull, "Emergency and on-demand health care: modelling a large complex system," *Journal of the Operational Research Society*, vol. 55, no. 1, pp. 34–42, 2004.

[17] P. Beraldi, M. E. Bruni, and D. Conforti, "Designing robust emergency medical service via stochastic programming," *European Journal of Operational Research*, vol. 158, no. 1, pp. 183–193, 2004.

[18] L. V. Green, S. Savin, and B. Wang, "Managing patient service in a diagnostic medical facility," *Operations Research*, vol. 54, no. 1, pp. 11–25, 2006.

[19] S. A. Jones, M. P. Joy, and J. Pearson, "Forecasting demand of emergency care," *Health Care Management Science*, vol. 5, no. 4, pp. 297–305, 2002.

[20] B. Friedman and M. Pauly, "Cost functions for a service firm with variable quality and stochastic demand: the case of hospitals," *The Review of Economics and Statistics*, vol. 63, no. 4, pp. 620–624, 1981.

[21] W. H. Wiersema, "Forecasting for inventory," *Electrical Apparatus*, vol. 12, article 40, no. 49, 1996.

[22] S. C. Graves, "A single-item inventory model for a non-stationary demand process," Working Papers WP, 3944-97, Massachusetts Institute of Technology (MIT). Sloan School of Management, 1997.

[23] P. Alfredsson and J. Verrijdt, "Modeling emergency supply flexibility in a two-echelon inventory system," *Management Science*, vol. 45, no. 10, pp. 1416–1431, 1999.

[24] E. S. Gardner Jr., "Evaluating forecast performance in an inventory control system," *Management Science*, vol. 36, no. 4, pp. 490–499, 1990.

[25] A. Federgruen and Y. Zheng, "Efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems," *Operations Research*, vol. 40, no. 4, pp. 808–813, 1992.

[26] C. Dance and A. A. Gaivoronski, "Stochastic optimization for real time service capacity allocation under random service demand," *Annals of Operations Research*, vol. 193, no. 1, pp. 221–253, 2012.

[27] J. Hwang, L. Gao, and W. Jang, "Joint demand and capacity management in a restaurant system," *European Journal of Operational Research*, vol. 207, no. 1, pp. 465–472, 2010.

[28] K. Gilbert, "An ARIMA supply chain model," *Management Science*, vol. 51, no. 2, pp. 305–310, 2005.

[29] J. D. Dana, "Using yield management to shift demand when the peak time is unknown," *RAND Journal of Economics*, vol. 30, no. 3, pp. 456–474, 1999.

[30] H. J. Finarelli Jr. and T. Johnson, "Effective demand forecasting in 9 steps," *Healthcare Financial Management*, vol. 58, no. 11, pp. 52–58, 2004.

[31] M. Utley, S. Gallivan, T. Treasure, and O. Valencia, "Analytical methods for calculating the capacity required to operate an effective booked admissions policy for elective inpatient services," *Health Care Management Science*, vol. 6, no. 2, pp. 97–104, 2003.

[32] M. Gaynor and W. B. Vogt, "Competition among hospitals," *RAND Journal of Economics*, vol. 34, no. 4, pp. 764–785, 2003.

[33] K. A. Ariyawansa, C. Cacho, and A. J. Felt, "A family of stochastic programming test problems based on a model for tactical manpower planning," *Journal of Mathematical Modelling and Algorithms*, vol. 4, no. 4, pp. 369–390, 2005.

[34] M. Armony, E. Plambeck, and S. Seshadri, "Sensitivity of optimal capacity to customer impatience in an unobservable M/M/S queue (why you shouldn't shout at the dmv)," *Manufacturing and Service Operations Management*, vol. 11, no. 1, pp. 19–32, 2009.

[35] J. Chomicki, D. Goldin, G. Kuper, and D. Toman, "Variable independence in constraint databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 6, pp. 1422–1436, 2003.

[36] N. Litvak, M. van Rijsbergen, R. J. Boucherie, and M. van Houdenhoven, "Managing the overflow of intensive care patients," *European Journal of Operational Research*, vol. 185, no. 3, pp. 998–1010, 2008.

[37] H. L. Lee and M. A. Cohen, "Multi-agent customer allocation in a stochastic service system," *Management Science*, vol. 31, no. 6, pp. 752–763, 1985.

[38] G. O. Roberts and J. S. Rosenthal, "Markov-chain Monte Carlo: some practical implications of theoretical results," *Canadian Journal of Statistics*, vol. 26, no. 1, pp. 5–29, 1998.

[39] M. P. Bailey, "Estimating service capacity using nonhomogeneous work arrivals," in *Proceedings of the 1993 Winter Simulation Conference*, pp. 408–413, ACM Press, 1993.

[40] D. Hughes and A. McGuire, "Stochastic demand, production responses and hospital costs," *Journal of Health Economics*, vol. 22, no. 6, pp. 999–1010, 2003.

[41] A. F. Seila, "On waiting times for a queue in which customers require simultaneous service from a random number of servers," *Operations Research*, vol. 32, no. 5, pp. 1181–1184, 1984.